

# Shared latent subspace modelling within Gaussian-Binary Restricted Boltzmann Machines for NIST i-Vector Challenge 2014

Danila Doroshin<sup>1</sup>, Alexander Yamshinin<sup>1</sup>, Nikolay Lubimov<sup>1</sup>,  
Marina Nastasenko<sup>1</sup>, Mikhail Kotov<sup>1</sup>, Maxim Tkachenko<sup>1</sup>

<sup>1</sup>Stel - Computer Systems Ltd., Moscow, Russia

{doroshin, yamshinin, lubimov, marina.nastasenko, kotov, tkachenko}@stel.ru

## Abstract

This paper presents a novel approach to speaker subspace modelling based on Gaussian-Binary Restricted Boltzmann Machines (GRBM). The proposed model is based on the idea of shared factors as in the Probabilistic Linear Discriminant Analysis (PLDA). GRBM hidden layer is divided into speaker and channel factors, herein the speaker factor is shared over all vectors of the speaker. Then Maximum Likelihood Parameter Estimation (MLE) for proposed model is introduced. Various new scoring techniques for speaker verification using GRBM are proposed. The results for NIST i-vector Challenge 2014 dataset are presented.

**Index Terms:** speaker recognition, speaker verification, Restricted Boltzmann Machines, i-vector, PLDA

## 1. Introduction

Actual approaches to text-independent automatic speaker verification (ASV) generally focus on the modelling of speaker and channel variability. The background of majority of these methods is based on factorising of the long-term distribution of spectral features. The standard method in ASV is to model this distribution using Gaussian Mixture Model (GMM) which is trained on a large audio database and referred as Universal Background Model (UBM). The Joint Factor Analysis technique [1] is based on decomposition of a UBM supervector<sup>1</sup> into the additive components belonging to speaker and channel subspace. Speaker and channel subspaces are modeled using low-dimensional factors. The i-vector approach is based on the total variability model [2] representing supervector in the low-dimensional space containing both speaker and channel information. Probabilistic Linear Discriminant Analysis (PLDA) [3] is applied to handle the influence of the channel variability in the i-vector space. PLDA deals with the decomposition of i-vectors on speaker and channel factors where the speaker factor is the same for all i-vectors of the speaker [4].

In this paper we examine an alternative way to effectively model speaker subspace using Restricted Boltzmann Machines (RBM). The idea is close to the PLDA factor modelling and based on dividing RBM hidden layer into the speaker and channel factors where the speaker factor is shared over all vectors of the speaker. The proposed model uses Gaussian-Binary RBM (GRBM) in contrast to a model described in [5] where Gaussian-Gaussian RBM was considered. The proposed approach is simply extended to the case of Binary-Binary RBM (BRBM). This choice is motivated by the ability of using Gaussian-Binary and Binary-Binary blocks as the internal

parts of deeper architectures as Deep Belief Networks and Deep Boltzmann Machines.

The paper is organized as follows. In section 2, the basic definitions of GRBMs are covered, then GRBM with shared latent subspace and corresponding generative model is introduced, MLE for proposed model including modification of contrastive divergence algorithm is performed. In section 2.4, various new scoring techniques for ASV are described including log-likelihood ratio (LLR) and normalized cosine scoring. In section 3 the train and test datasets are described. The results for NIST i-vector Challenge 2014 dataset are given and compared to the baseline and state-of-the-art methods. In section 4, conclusions and future work directions are discussed. In the appendix section 5 some theoretical proofs are presented.

## 2. Shared latent subspace modelling within Gaussian-Binary Restricted Boltzmann Machines

### 2.1. General GRBM

GRBM defines probability density function (PDF) with input visible variable  $x$  and hidden (latent) variable  $h$  [6, 7]

$$P(x, h) = \frac{1}{Z} e^{-E(x, h)}$$

where  $Z$  is a normalizing constant called a partition function and  $E(x, h)$  is an energy function. For GRBM  $x$  is from the continuous space  $\mathbb{R}^p$  and  $h$  is from the discrete space  $h \in \{0, 1\}^r$ .  $E(x, h)$  depends on visible bias  $b$ , hidden bias  $d$ , vector of standard deviations  $\sigma$  and connectivity matrix  $W$

$$E(x, h) = \frac{1}{2} \left\| \frac{x - b}{\sigma} \right\|^2 - d^T h - \left( \frac{x}{\sigma^2} \right)^T W h$$

Here and then  $\star/\star$  denotes element-wise division of vectors,  $\star^2$  denotes element-wise squaring,  $T$  denotes transposition and  $\|\star\|$  is Euclidean norm.

### 2.2. GRBM with shared latent subspace

GRBM is modified to simulate speaker and channel variability. The hidden variable is divided into the speaker factor  $s$  and the channel factor  $c$ , i.e.  $h = [s; c]$ . According to this, parameters are split into two groups, i.e.  $d = [f; g]$ ,  $W = [F, G]$ . Rewrite the energy function expression using split parameters

$$E(x, s, c) = \frac{1}{2} \left\| \frac{x - b}{\sigma} \right\|^2 - f^T s - g^T c - \left( \frac{x}{\sigma^2} \right)^T (Fs + Gc)$$

The speaker factor is supposed to be the same for all i-vectors of one speaker while the channel factors are individual for each

<sup>1</sup>A supervector is a vector of stacked GMM mean vectors

i-vector. Below a set of generative models depending on the number of i-vectors corresponding to the speaker and PDF for them are introduced. Consider the case of  $N$  i-vectors of the speaker and correspondent  $N$ -order generative model. Denote speaker data as  $X = \{x_1, x_2, \dots, x_N\}$ , channel factors as  $C = \{c_1, c_2, \dots, c_N\}$  and speaker factor as  $s$ . PDF for  $N$ -order model is expressed as follows

$$P_N(X, s, C) = \frac{1}{Z_N} e^{-E_N(X, s, C)} \quad (1)$$

where  $E_N(X, s, C) = \sum_{n=1}^N E(x_n, s, c_n)$  and  $Z_N = \int_X \sum_{s, C} e^{-E_N(X, s, C)} dX$ . The generative process for this model is shown in Figure 1. First  $s, C$  are generating accord-

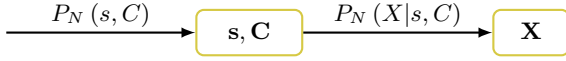


Figure 1: The generative process for  $N$ -order model

ing the distribution  $P_N(s, C) = \int P_N(X, s, C) dX$  then  $X$  is generating according the distribution  $P_N(X|s, C)$ , where

$$P_N(X|s, C) = \prod_{n=1}^N \mathcal{N}(x_n, b + Fs + Gc_n, \sigma^2) \quad (2)$$

and  $\mathcal{N}$  denotes Gaussian distribution.

### 2.3. Maximum likelihood parameter estimation

Assume we have a labeled training set of  $K$  speakers, denoted by  $\mathbb{X} = \{X_k\}_{k=1}^K$  where  $X_k$  is data with  $N_k$  i-vectors that corresponds to  $k$ -th speaker. Let  $N_k \in \{2, 3, \dots, M\}$ , hence there are  $M - 1$  generative models, and it is assumed that their parameters are tied. The aim is to estimate the set of parameters  $\Theta = \{f, g, F, G, b, \sigma\}$  using MLE criterium that is standard approach for RBMs [7]. For the optimization of MLE objective function we use a stochastic gradient descent approach that is widely used for RBMs [8, 6]. Since data for each pair of speakers are assumed to be independent, normalized log-likelihood function takes the form of sum of log-likelihood functions for each generative model

$$\begin{aligned} \mathcal{L}_{norm}(\mathbb{X}|\Theta) &= \frac{1}{\sum_k N_k} \sum_k \mathcal{L}(X_k|\Theta) = \\ &= \frac{1}{\sum_k N_k} \sum_{N=2}^M \sum_{k: N_k=N} \mathcal{L}(X_k|\Theta) \end{aligned}$$

Further speaker's index will be neglected and there will be discussed the likelihood of the data from one speaker. Denote speaker data as  $X = \{x_1, x_2, \dots, x_N\}$  then

$$\mathcal{L}(X|\Theta) = \log P_N(X) \quad (3)$$

Denote the realization of speaker factor as  $s$  and the realizations of channel factors as  $c_n$ ,  $C = \{c_1, c_2, \dots, c_N\}$ . Consider log-likelihood from (3) marginalizing (1) over all possible values of latent variables

$$\mathcal{L}(X|\Theta) = \log \sum_{s, C} e^{-E_N(X, s, C)} - \log Z_N \quad (4)$$

Consider the first part of gradient of (4), making the same transformations as for the general GRBM [7]

$$\begin{aligned} \nabla_{\Theta} \mathcal{L}_1(X|\Theta) &= - \frac{\sum_{s, C} P_N(X, s, C) \frac{\partial}{\partial \Theta} E_N(X, s, C)}{P_N(X)} = \\ &= - \sum_s \sum_{n=1}^N \sum_{c_n} P_N(s, c_n|X) \frac{\partial E(x_n, s, c_n)}{\partial \Theta} \end{aligned}$$

As a result, the gradient of (4) is represented as the following sum

$$\nabla_{\Theta} \mathcal{L}(X|\Theta) = \nabla_{\Theta} \mathcal{L}_1(X|\Theta) - \mathcal{E}_{P_N(\tilde{X})} [\nabla_{\Theta} \mathcal{L}_1(\tilde{X}|\Theta)] \quad (5)$$

Here  $\mathcal{E}$  denotes expectation:  $\mathcal{E}_{P_N(\tilde{X})} [\star] = \int_{\tilde{X}} P_N(\tilde{X}) \star d\tilde{X}$ . The modification of the contrastive divergence algorithm [7] that enables to compute the second term of the gradient (5) is presented in section 2.3.1. Below the gradient of the first term will be considered. Taking into account the derivatives of energy function [8] the gradient of  $\mathcal{L}_1(X|\Theta)$  takes the following form

$$\nabla_{F_{ij}} \mathcal{L}_1(X|\Theta) = P_N(s_j = 1|X) \frac{\bar{x}_i}{\sigma_i^2} \quad (6)$$

$$\nabla_{f_i} \mathcal{L}_1(X|\Theta) = N P_N(s_j = 1|X) \quad (7)$$

$$\nabla_{G_{ij}} \mathcal{L}_1(X|\Theta) = \sum_n P_N(c_{nj} = 1|X) \frac{x_{ni}}{\sigma_i^2} \quad (8)$$

$$\nabla_{g_i} \mathcal{L}_1(X|\Theta) = \sum_n P_N(c_{nj} = 1|X) \quad (9)$$

$$\nabla_{b_i} \mathcal{L}(X|\Theta) = \frac{1}{\sigma_i^2} (\bar{x}_i - N b_i) \quad (10)$$

$$\nabla_{z_i} \mathcal{L}_1(X|\Theta) = - \frac{\bar{x}_i}{\sigma_i^2} \sum_j F_{ij} P_N(s_j = 1|X) - \quad (11)$$

$$- \sum_{n,j} \frac{x_{ni}}{\sigma_i^2} G_{ij} P_N(c_{nj} = 1|X) + \frac{1}{2} \sum_n \frac{(x_{ni} - b_i)^2}{\sigma_i^2}$$

Here and further  $\bar{x} = \sum_{n=1}^N x_n$  and  $i, j$  denote indexing over dimensions. Additionally, instead of  $\sigma_i$ , we update log-variances  $z_i = \log \sigma_i^2$  which are naturally constrained to stay positive [9]. Posterior probabilities of latent factors from the expressions (6-11) are determined from the following relations, which are proved in the appendix section of the paper

$$P_N(s_j = 1|X) = \text{sigm} \left( N f_j + (\bar{x}/\sigma^2)^T F_{*j} \right) \quad (12)$$

$$P_N(c_{nj} = 1|X) = \text{sigm} \left( g_j + (x_n/\sigma^2)^T G_{*j} \right) \quad (13)$$

Here and further  $F_{*j}$  and  $G_{*j}$  denotes respectively  $j$ -th column of the matrices and  $\text{sigm}(\star) = 1/(1 + e^{-\star})$ . From the expression (13) it is clear that posterior for  $c_n$  depends only on  $x_n$  and it is the same as for the general GRBM. The main difference is that all speaker's i-vectors  $X$  influence the speaker factor posterior (12).

#### 2.3.1. Contrastive divergence

The modification of the contrastive divergence algorithm [7] is presented below. It enables to compute approximately the second part of the gradient (5). Expectation is replaced by mean over a finite set of samples from distribution  $P_N(\tilde{X})$ . Since it is hard to get these samples because of the complexity of the generative process, an approximate algorithm called the m-steps

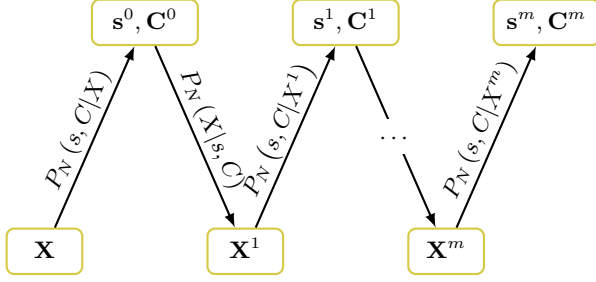


Figure 2: Contrastive divergence

contrastive divergence [10] is applied. Algorithm scheme is presented in Figure 2. Data of the speaker  $X$  is used to initialize the algorithm on the zero step. Intermediate  $k$ -th step of the algorithm is presented below. Reconstruction of visible data  $X^k$  is sampled using (2). Latent variables  $s^k, C^k$  are sampled using (12) and (13). For binarization we use uniformly distributed random thresholds following recommendations from [7].

#### 2.4. Scoring

In this section various scoring strategies for GRBM with shared latent subspace will be presented.

##### 2.4.1. Log-likelihood scoring

The LLR for a given verification trial  $\{X, x_t\}$ , i.e.  $X$  is a set of  $N$  enrollment speaker's vectors and  $x_t$  is a test vector, is the LLR between target and non-target hypotheses. The target hypothesis is that the trial vectors share a common speaker factor, i.e. generated by  $N + 1$ -order model. Non-target hypothesis is that  $X$  is generated by  $N$ -order model and  $x_t$  is independent from them and generated by 1-order model.

$$l = \log \frac{P_{N+1}(X, x_t)}{P_N(X)P_1(x_t)}$$

The expression for the LLR score is given below and its proof is given in the appendix section

$$l = \sum_i \log \frac{1 + e^{(N+1)f_i + ((\bar{x} + x_t)/\sigma^2)^T F_{*i}}}{\left(1 + e^{Nf_i + (\bar{x}/\sigma^2)^T F_{*i}}\right) \left(1 + e^{f_i + (x_t/\sigma^2)^T F_{*i}}\right)} + \log \frac{Z_N Z_1}{Z_{N+1}}$$

Some methods exist for the approximate computation of the partition function [11]. Note that values of the partition function do not influence the performance of the system in case when all speakers have the same number of enrollment vectors.

##### 2.4.2. Cosine scoring

We apply the standard cosine scoring [12] to i-vectors previously projected onto the subspace  $F^T$ . Denote  $y_n = \frac{F^T x_n}{\|F^T x_n\|}$  for each speaker's i-vector from  $X$  and  $y_t = \frac{F^T x_t}{\|F^T x_t\|}$  for test. The score is cosine between average speaker's vector  $y_{sp} = \sum_n y_n / N$  and the test vector

$$l_{cos} = y_t^T \frac{y_{sp}}{\|y_{sp}\|}$$

In addition to the general cosine score, we propose normalized cosine score  $l_{norm}$ . It takes into account information on the

width of the speaker's cluster that is lost in the standard cosine scoring. General cosine score is divided by the average cosine within the speaker's set  $cos_{sp} = \sum_n y_n^T \frac{y_{sp}}{\|y_{sp}\|} / N$ . It can be shown that  $cos_{sp} = \|y_{sp}\|$ . Taking it into account, the expression for the normalized cosine score takes the form

$$l_{norm} = \frac{l_{cos}}{\|y_{sp}\|}$$

##### 2.4.3. PLDA on F-projected i-vectors

PLDA model is trained on i-vectors projected onto the subspace  $F^T$  and then projected on unit sphere –  $y_n$ . PLDA handles residual channel variability using linear factor model [3]. Scoring is done using the LLR for PLDA model [13, 14].

### 3. Experimental results

#### 3.1. Dataset

*NIST i-vector Machine Learning Challenge 2014* dataset has been chosen to test the efficiency of the proposed model. The dataset consists of a labeled development set (*devset*), a labeled model set (*modelset*) with 5 i-vectors per model and an unlabeled test set (*testset*). Since labels for the *devset* were not available during the challenge, the best results were obtained from methods that allowed to cluster the *devset* and then to apply PLDA [15, 16].

In our experiments we reformed the dataset. Preliminary all i-vectors with duration less than 10 seconds have been removed for their bad quality [15, 16]. We construct a new labeled *trainset*, *modelset*, *testset*, *modelsetCV*, *testsetCV*. Speakers from *devset* with 3 to 10 i-vectors united with the initial *modelset* are assigned to the *trainset*, with 11 to 15 i-vectors are assigned to the new *modelset* and *testset*, remaining speakers with more than 15 i-vectors form cross validation set (*modelsetCV*, *testsetCV*). First 5 i-vectors from each speaker's set form enrollment in the *modelset* and the remaining form the *testset*. The same is done for the cross validation set. Eventually the *trainset* contains 3281 speakers and total 18759 i-vectors, 717 speakers with 3585 i-vectors and 5400 i-vectors in the *modelset* and the *testset* respectively. We used minDCF as a measure of the system performance and a measure for the cross validation processing

$$minDCF = \min_{th} FR(th) + 100FA(th)$$

where *FA* and *FR* denote the false acceptance and the false rejection rates, and *th* the varying threshold. The trials consist of all possible pairs involving a target speaker set from the *modelset* and a test i-vector from the *testset*.

#### 3.2. Parameters estimation

Whitened [17] *trainset* is used for the parameter estimation. The parameters of whitening are computed on the *trainset* too. This transform is used further for all trials. We set initial biases  $f$ ,  $g$  and  $b$  to the zero. Following the recommendations from [7] elements of the connectivity matrices  $F$  and  $G$  are generated using normal distribution with zero mean and standard deviation equal to 0.01. Elements of standard deviation vector  $\sigma$  were set to 1.0. The case of  $\sigma$  reestimation showed the worse results.

The best performance was obtained using the speaker factor dimension equal to 500, the channel factor dimension equal to 100 while i-vector dimension equal to 600. We used the mini-batch stochastic gradient descent algorithm [7] with learning

rate 0.01, momentum 0.5 and zero weight decay. Each batch contained 256 speakers. After each epoch the speakers are shuffled between batches. It took 40 epochs to achieve the best minDCF on the cross validation set. In case when all speakers belong to one batch it took 10 times more iterations to reach the same performance of the system. To train PLDA model on i-vectors, whitened *trainset* was projected on the unit sphere [17]. It was found that the best speaker and channel factor dimensions for PLDA are equal to 590 and 10 respectively. PLDA model trained on i-vectors that were projected on  $F^T$  has the speaker and channel factor dimensions equal to 499 and 1 respectively. Increase of the channel factor dimension showed the worse results.

### 3.3. Results

We compare our algorithm with the NIST 2014 baseline cosine scoring and the state of the art [15, 16] PLDA. As the results

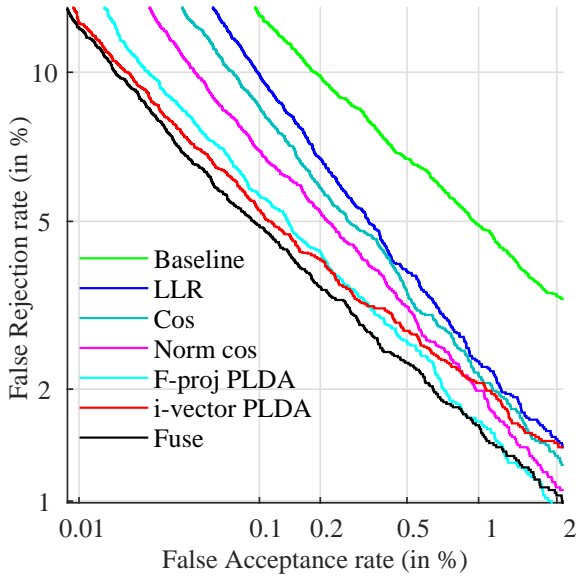


Figure 3: Comparison of the proposed scoring algorithms with NIST 2014 baseline cosine and PLDA

	Baseline	LLR	Cos
EER (in %)	2.81	1.68	1.58
minDCF	0.210	0.185	0.167
Norm cos	F-proj PLDA	i-vector PLDA	Fuse
1.43	<b>1.30</b>	1.51	1.33
0.145	0.123	<b>0.114</b>	<b>0.108</b>

Table 1: Results on NIST 2014 dataset.

in Table 1 and Figure 3 demonstrate, all scoring strategies perform better than challenge baseline. Despite the optimality of log-likelihood GRBM scoring, it did not show the best results among the other GRBM scoring strategies. Perhaps, this is due to the specific of the i-vector data. The considered normalized cosine scoring performs better than the standard cosine scoring. In terms of EER, the best result is achieved on PLDA trained

on i-vectors projected on speaker space  $F^T$  of GRBM. In addition, linear fuse [18] of two PLDA models is presented. The first model uses i-vectors as features and the second one uses i-vectors projected on  $F^T$ . Coefficients of the fuse were estimated on the cross validation set by using logistic regression training with weighted MLE criterium [19]. As can be seen in Figure 3, the fused scores outperform i-vector PLDA in the area of low FA and retain performance in the EER area.

## 4. Conclusions and Further Work

We used shared latent subspace in GRBM hidden layer to separate speaker dependent and speaker independent factors in i-vector space. Approximate maximum likelihood parameters estimation is presented. For the proposed model several scoring methods for the speaker verification were considered, including a novel log-likelihood scoring and normalized cosine scoring. PLDA operating with i-vectors projected on GRBM speaker space performed results that are comparable to the state of the art i-vector PLDA approach. Fuse of these two PLDA models showed the best results at all operating points.

In further work, the method of projection on GRBM speaker space can be viewed as a stand-alone channel variability compensation technique. GRBM with shared latent subspace can be extended to the other types of RBM and can be used as a block in deeper architectures.

## 5. Appendix

In this section proofs of LLR score expression from section 2.4 and expressions (12), (13) are derived. They can be obtained if there is an expression for a posterior probability  $P_N(s, C|X)$ . First we derive joint PDF for latent variables and data  $P_N(s, C, X)$  using its definition (1)

$$P_N(s, C, X) = \mathcal{C}_{N,X} \cdot \quad (14)$$

$$\prod_{i,n,j} e^{N f_i s_i + (\bar{x}/\sigma^2)^T F_{*i} s_i} e^{g_j c_{nj} + (x_n/\sigma^2)^T G_{*j} c_{nj}}$$

Here  $\mathcal{C}_{N,X} = \frac{1}{Z_N} e^{-\frac{1}{2} \sum_n \|\frac{x_n - b}{\sigma}\|^2}$ . Marginalizing (14) over all possible latent variables we have

$$P_N(X) = \mathcal{C}_{N,X} \cdot \quad (15)$$

$$\prod_{i,n,j} \left( 1 + e^{N f_i + (\bar{x}/\sigma^2)^T F_{*i}} \right) \left( 1 + e^{g_j + (x_n/\sigma^2)^T G_{*j}} \right)$$

Eventually the posterior probability of latent variables is the division of (14) on (15)

$$P_N(s, C|X) = \quad (16)$$

$$\prod_i \frac{e^{N f_i s_i + (\bar{x}/\sigma^2)^T F_{*i} s_i}}{1 + e^{N f_i + (\bar{x}/\sigma^2)^T F_{*i}}} \prod_{n,j} \frac{e^{g_j c_{nj} + (x_n/\sigma^2)^T G_{*j} c_{nj}}}{1 + e^{g_j + (x_n/\sigma^2)^T G_{*j}}}$$

Now expressions for (12), (13) can be obtained by summing (16) over corresponding latent variables. The expression for LLR score is obtained by applying (15) to the three trial subsets.

## 6. Acknowledgement

Research is conducted by Stel - Computer systems Ltd. with support of the Ministry of Education and Science of the Russian Federation (Contract 14.579.21.0058) Unique ID for Applied Scientific Research (project) RFMEFI57914X0058. The data presented, the statements made, and the views expressed are solely the responsibility of the authors.

## 7. References

- [1] P. Kenny, "Joint factor analysis of speaker and session variability: Theory and algorithms," *CRIM, Montreal, (Report) CRIM-06/08-13*, 2005.
- [2] N. Dehak, P. Kenny, R. Dehak, P. Dumouchel, and P. Ouellet, "Front-end factor analysis for speaker verification," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 19, no. 4, pp. 788–798, 2011.
- [3] S. J. Prince and J. H. Elder, "Probabilistic linear discriminant analysis for inferences about identity," in *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*. IEEE, 2007, pp. 1–8.
- [4] P. Kenny, P. Ouellet, N. Dehak, V. Gupta, and P. Dumouchel, "A study of interspeaker variability in speaker verification," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 16, no. 5, pp. 980–988, 2008.
- [5] T. Stafylakis, P. Kenny, M. Senoussaoui, and P. Dumouchel, "Plda using gaussian restricted boltzmann machines with application to speaker verification," in *INTERSPEECH*, 2012.
- [6] R. Salakhutdinov, "Learning deep generative models," Ph.D. dissertation, University of Toronto, 2009.
- [7] G. Hinton, "A practical guide to training restricted boltzmann machines," *Momentum*, vol. 9, no. 1, p. 926, 2010.
- [8] N. Wang, J. Melchior, and L. Wiskott, "Gaussian-binary restricted boltzmann machines on modeling natural image statistics," *arXiv preprint arXiv:1401.5900*, 2014.
- [9] K. Cho, A. Ilin, and T. Raiko, "Improved learning of gaussian-bernoulli restricted boltzmann machines," in *Artificial Neural Networks and Machine Learning–ICANN 2011*. Springer, 2011, pp. 10–17.
- [10] G. Hinton, "Training products of experts by minimizing contrastive divergence," *Neural computation*, vol. 14, no. 8, pp. 1771–1800, 2002.
- [11] Y. Burda, R. B. Grosse, and R. Salakhutdinov, "Accurate and conservative estimates of mrf log-likelihood using reverse annealing," *arXiv preprint arXiv:1412.8566*, 2014.
- [12] M. Senoussaoui, P. Kenny, N. Dehak, and P. Dumouchel, "An i-vector extractor suitable for speaker recognition with both microphone and telephone speech," in *Odyssey*, 2010, p. 6.
- [13] A. Larcher, K. A. Lee, B. Ma, and H. Li, "Phonetically-constrained plda modeling for text-dependent speaker verification with multiple short utterances," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*. IEEE, 2013, pp. 7673–7677.
- [14] P. Rajan, A. Afanasyev, V. Hautamäki, and T. Kinnunen, "From single to multiple enrollment i-vectors: Practical plda scoring variants for speaker verification," *Digital Signal Processing*, vol. 31, pp. 93–101, 2014.
- [15] E. Khoury, L. El Shafey, M. Ferras, and S. Marcel, "Hierarchical speaker clustering methods for the nist i-vector challenge," in *Odyssey: The Speaker and Language Recognition Workshop*, no. EPFL-CONF-198439, 2014.
- [16] S. Novoselov, T. Pekhovsky, and K. Simonchik, "Stc speaker recognition system for the nist i-vector challenge," in *Odyssey: The Speaker and Language Recognition Workshop*, 2014, pp. 231–240.
- [17] D. Garcia-Romero and C. Y. Espy-Wilson, "Analysis of i-vector length normalization in speaker recognition systems," in *Inter-speech*, 2011, pp. 249–252.
- [18] N. Brummer, L. Burget, J. H. Cernocky, O. Glembek, F. Grezl, M. Karafiat, D. A. Van Leeuwen, P. Matejka, P. Schwarz, and A. Strasheim, "Fusion of heterogeneous speaker recognition systems in the stbu submission for the nist speaker recognition evaluation 2006," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 15, no. 7, pp. 2072–2084, 2007.
- [19] D. A. van Leeuwen and N. Brümmer, "The distribution of calibrated likelihood-ratios in speaker recognition," *arXiv preprint arXiv:1304.1199*, 2013.